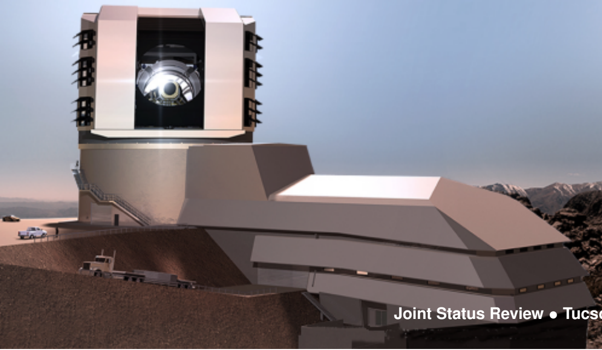
# Data Management Breakout: Management

William O'Mullane, AURA/LSST
DM Project Manager

NSF/DOE Joint Status Review
28th Aug 2019



# Outline

Introduction

Estimate to Complete & Current Staffing (Charge 2, 3)

Progress tracking (Charge 1, 2, 3)

Risks and Opportunities (Charge 1, 3)

Previous Recomendations (Charge 7)

Conclusion

# Introduction

- We shall cover ETC for each top level WBS element
- We shall look at some of the non-EVMS progress tracking
- We shall review recommendations from the previous reviews

# Outline

# DM Budget Summary

CHARGE: 2

| | | CUMULATIVE TO DATE ($k) 31 May 2019 | | | | | | | AT COMPLETION | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | BUDGETED COST | | ACTUAL | VARIANCE | | PERFORMANCE | | | | |
| WBS | DESCRIPTION | WORK SCHEDULED | WORK PERFORMED | COST OF WORK PERFORMED | SCHEDULE | COST | SPI | CPI | BAC | EAC | % COMP |
| 1.02C | Data Management Construction | $ 88,798 | $ 86,292 | $ 86,010 | $ (2,505) | $ 282 | 0.97 | 1.00 | $ 147,218 | $ 148,124 | 59% |
| 1.02C.01 | System Management | $ 5,203 | $ 5,203 | $ 5,082 | $ - | $ 121 | 1.00 | 1.02 | $ 6,938 | $ 6,970 | 75% |
| 1.02C.02 | System Engineering | $ 6,034 | $ 6,034 | $ 6,339 | $ - | $ (305) | 1.00 | 0.95 | $ 9,496 | $ 9,850 | 64% |
| 1.02C.03 | Alert Production | $ 6,163 | $ 6,025 | $ 6,116 | $ (138) | $ (90) | 0.98 | 0.99 | $ 11,495 | $ 11,685 | 52% |
| 1.02C.04 | Data Release Production | $ 8,194 | $ 7,640 | $ 7,024 | $ (554) | $ 616 | 0.93 | 1.09 | $ 17,493 | $ 17,311 | 44% |
| 1.02C.05 | Science User Interface and Analysis Tools | $ 7,627 | $ 7,499 | $ 7,362 | $ (128) | $ 138 | 0.98 | 1.02 | $ 8,667 | $ 8,577 | 87% |
| 1.02C.06 | Science Data Archive and Application Services | $ 9,409 | $ 8,841 | $ 8,983 | $ (568) | $ (143) | 0.94 | 0.98 | $ 13,052 | $ 13,400 | 68% |
| 1.02C.07 | Processing Control and Site Infrastructure | $ 18,401 | $ 17,942 | $ 17,823 | $ (460) | $ 118 | 0.98 | 1.01 | $ 42,122 | $ 42,218 | 43% |
| 1.02C.08 | International Communications and Base Site | $ 24,400 | $ 23,826 | $ 24,059 | $ (575) | $ (233) | 0.98 | 0.99 | $ 30,487 | $ 30,686 | 78% |
| 1.02C.09 | Data Management System Integration and Te: | $ - | $ - | $ 3 | $ - | $ (3) | 0.00 | - | $ 815 | $ 827 | 0% |
| 1.02C.10 | Science Quality and Reliability Engineering | $ 3,366 | $ 3,282 | $ 3,221 | $ (83) | $ 61 | 0.98 | 1.02 | $ 6,654 | $ 6,599 | 49% |

- Current BAC ∼$147M
- Corresponds to ∼ 400 Person Years and $55M non-labor costs
- As of May 2019, ∼65 FTE provided by 96 individuals

# Management, Science and Sys Eng

**CHARGE:** 2, 3

| | | CUMULATIVE TO DATE ($k) 31 May 2019 | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | BUDGETED COST | | ACTUAL | VARIANCE | | PERFORMANCE | | AT COMPLETION | | |
| WBS | DESCRIPTION | WORK SCHEDULED | WORK PERFORMED | COST OF WORK PERFORMED | SCHEDULE | COST | SPI | CPI | BAC | EAC | % COMP |
| **1.02C** | **Data Management Construction** | **$ 10,288** | **$ 10,288** | **$ 10,510** | **$ -** | **$ (223)** | **1.00** | **0.98** | **$ 16,300** | **$ 16,734** | **63%** |
| **1.02C.01** | **System Management** | **$ 4,254** | **$ 4,254** | **$ 4,169** | **$ -** | **$ 85** | **1.00** | **1.02** | **$ 5,989** | **$ 6,057** | **71%** |
| 1.02C.01.01 | Project Management | $ 4,254 | $ 4,254 | $ 4,169 | $ - | $ 85 | 1.00 | 1.02 | $ 5,989 | $ 6,057 | 71% |
| **1.02C.02** | **System Engineering** | **$ 6,034** | **$ 6,034** | **$ 6,339** | **$ -** | **$ (305)** | **1.00** | **0.95** | **$ 9,496** | **$ 9,850** | **64%** |
| 1.02C.02.01 | Data Management Science | $ 2,517 | $ 2,517 | $ 2,703 | $ - | $ (186) | 1.00 | 0.93 | $ 3,812 | $ 3,999 | 66% |
| 1.02C.02.02 | Data Management Architecture | $ 3,516 | $ 3,516 | $ 3,635 | $ - | $ (119) | 1.00 | 0.97 | $ 5,684 | $ 5,851 | 62% |
| **1.02C.09** | **Data Management System Integration and Test** | **$ -** | **$ -** | **$ 3** | **$ -** | **$ (3)** | **-** | **-** | **$ 815** | **$ 827** | **0%** |
| 1.02C.09.01 | Archive Center Integration | $ - | $ - | $ - | $ - | $ - | - | - | $ 108 | $ 110 | 0% |
| 1.02C.09.02 | Base Center Integration | $ - | $ - | $ 3 | $ - | $ (3) | - | - | $ - | $ 3 | 0% |
| 1.02C.09.04 | US Data Access Center Integration | $ - | $ - | $ - | $ - | $ - | - | - | $ 707 | $ 714 | 0% |

- System Management (02C.01.01)
  - 1 Project Manager; 1 FTE
- DM Science (02C.02.01)
  - 5 Scientists; 3.6 FTE
- DM Architecture (02C.02.02)
  - 5 Engineers; 4.1 FTE

Most work is LOE.

# Science Pipelines I

CHARGE: 2

| | | CUMULATIVE TO DATE ($k) 31 May 2019 | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | BUDGETED COST | | ACTUAL | VARIANCE | | PERFORMANCE | | AT COMPLETION | | |
| WBS | DESCRIPTION | WORK SCHEDULED | WORK PERFORMED | COST OF WORK PERFORMED | SCHEDULE | COST | SPI | CPI | BAC | EAC | % COMP |
| **1.02C** | **Data Management Construction** | **$ 14,357** | **$ 13,666** | **$ 13,139** | **$ (692)** | **$ 526** | **0.95** | **1.04** | **$ 28,988** | **$ 28,996** | **47%** |
| **1.02C.03** | **Alert Production** | **$ 6,163** | **$ 6,025** | **$ 6,116** | **$ (138)** | **$ (90)** | **0.98** | **0.99** | **$ 11,495** | **$ 11,685** | **52%** |
| 1.02C.03.00 | Management, Leadership & Other Costs | $ 3,220 | $ 3,220 | $ 3,139 | $ - | $ 81 | 1.00 | 1.03 | $ 3,923 | $ 3,842 | 82% |
| 1.02C.03.01 | Single Frame Processing | $ 340 | $ 340 | $ 285 | $ - | $ 55 | 1.00 | 1.19 | $ 1,489 | $ 1,434 | 23% |
| 1.02C.03.02 | Catalog Association for Alert Production | $ 103 | $ 94 | $ 87 | $ (8) | $ 7 | 0.92 | 1.08 | $ 251 | $ 244 | 38% |
| 1.02C.03.03 | Alert Distribution System | $ 222 | $ 222 | $ 262 | $ - | $ (39) | 1.00 | 0.85 | $ 659 | $ 698 | 34% |
| 1.02C.03.04 | Alert Generation Pipeline | $ 751 | $ 691 | $ 774 | $ (60) | $ (84) | 0.92 | 0.89 | $ 1,208 | $ 1,315 | 57% |
| 1.02C.03.05 | Tools for Science Pipelines | $ 1,009 | $ 989 | $ 819 | $ (20) | $ 170 | 0.98 | 1.21 | $ 1,796 | $ 1,664 | 55% |
| 1.02C.03.06 | Moving Object Processing System (MOPS) | $ 152 | $ 106 | $ 147 | $ (46) | $ (40) | 0.70 | 0.72 | $ 854 | $ 926 | 12% |
| 1.02C.03.07 | Transform Fitting on Stacks of Images | $ 124 | $ 120 | $ 234 | $ (4) | $ (114) | 0.97 | 0.51 | $ 738 | $ 857 | 16% |
| 1.02C.03.08 | Integration | $ 244 | $ 244 | $ 369 | $ - | $ (125) | 1.00 | 0.66 | $ 578 | $ 703 | 42% |
| **1.02C.04** | **Data Release Production** | **$ 8,194** | **$ 7,640** | **$ 7,024** | **$ (554)** | **$ 616** | **0.93** | **1.09** | **$ 17,493** | **$ 17,311** | **44%** |
| 1.02C.04.00 | Management, Leadership & Other Costs | $ 3,463 | $ 3,463 | $ 3,380 | $ - | $ 83 | 1.00 | 1.02 | $ 3,959 | $ 4,207 | 87% |
| 1.02C.04.01 | Software Primitives | $ 1,159 | $ 1,078 | $ 1,093 | $ (81) | $ (15) | 0.93 | 0.99 | $ 2,258 | $ 2,293 | 48% |
| 1.02C.04.02 | Calibration Products | $ 807 | $ 702 | $ 495 | $ (105) | $ 207 | 0.87 | 1.42 | $ 1,903 | $ 1,713 | 37% |
| 1.02C.04.03 | Image Characterization | $ 372 | $ 286 | $ 280 | $ (86) | $ 7 | 0.77 | 1.02 | $ 953 | $ 951 | 30% |
| 1.02C.04.04 | Coaddition | $ 316 | $ 231 | $ 267 | $ (85) | $ (36) | 0.73 | 0.86 | $ 1,474 | $ 1,513 | 16% |
| 1.02C.04.05 | Detection & Deblending | $ 442 | $ 411 | $ 295 | $ (31) | $ 116 | 0.93 | 1.39 | $ 2,328 | $ 2,217 | 18% |
| 1.02C.04.06 | Characterization & Measurement | $ 998 | $ 921 | $ 928 | $ (77) | $ (7) | 0.92 | 0.99 | $ 3,104 | $ 3,121 | 30% |
| 1.02C.04.07 | Maintenance, Quality & Documentation | $ 637 | $ 548 | $ 286 | $ (89) | $ 262 | 0.86 | 1.91 | $ 1,514 | $ 1,296 | 36% |

# Science Pipelines II

CHARGE: 3

– Alert Production (02C.03)

- 11 persons; 9.6 FTE
- 1 Science Lead; 1 T/CAM; 8 Scientists/Engineers
- Some staff may be temporarily reassigned to the Telescope & Site Subsystem during mid-late 2019
- Recruitment starting for a new engineer to work on the alert distribution system.

– Data Release Production (02C.04)

- 15 persons; 12 FTE
- 1 Science Lead; 0.5 T/CAM; 11.5 Scientists/Engineers
- Increasing to 18 persons/15 FTE over the next six months

# Science User Interface & Tools (02C.05)

**CHARGE:** 2, 3

| | | CUMULATIVE TO DATE ($k) 31 May 2019 | | | | | | | AT COMPLETION | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | BUDGETED COST | | ACTUAL | VARIANCE | | PERFORMANCE | | | | |
| WBS | DESCRIPTION | WORK SCHEDULED | WORK PERFORMED | COST OF WORK PERFORMED | SCHEDULE | COST | SPI | CPI | BAC | EAC | % COMP |
| 1.02C | Data Management Construction | $ 7,627 | $ 7,499 | $ 7,362 | $ (128) | $ 138 | 0.98 | 1.02 | $ 8,667 | $ 8,577 | 87% |
| 1.02C.05 | Science User Interface and Analysis Tools | $ 7,627 | $ 7,499 | $ 7,362 | $ (128) | $ 138 | 0.98 | 1.02 | $ 8,667 | $ 8,577 | 87% |
| 1.02C.05.00 | Management, Leadership & Other Costs | $ 2,673 | $ 2,673 | $ 2,681 | $ - | $ (8) | 1.00 | 1.00 | $ 2,965 | $ 2,986 | 90% |
| 1.02C.05.01 | Basic Archive Access Tools | $ 2,317 | $ 2,294 | $ 2,301 | $ (23) | $ (8) | 0.99 | 1.00 | $ 2,317 | $ 2,318 | 99% |
| 1.02C.05.02 | Data Analysis and Visualization Tools | $ 1,161 | $ 1,161 | $ 1,132 | $ - | $ 30 | 1.00 | 1.03 | $ 1,161 | $ 1,132 | 100% |
| 1.02C.05.03 | Alert/Notification Toolkit | $ 27 | $ 27 | $ 11 | $ - | $ 15 | 1.00 | 2.34 | $ 27 | $ 11 | 100% |
| 1.02C.05.04 | User Assistance/Help Desk | $ 2 | $ 2 | $ - | $ - | $ 2 | 1.00 | - | $ 2 | $ - | 100% |
| 1.02C.05.05 | User Workspace Toolkit | $ 141 | $ 141 | $ 140 | $ - | $ 1 | 1.00 | 1.00 | $ 141 | $ 140 | 100% |
| 1.02C.05.06 | Client-server Query & Visualization Framewor | $ 530 | $ 514 | $ 531 | $ (16) | $ (16) | 0.97 | 0.97 | $ 530 | $ 545 | 97% |
| 1.02C.05.07 | LSST Science Platform Interfaces | $ 265 | $ 236 | $ 211 | $ (29) | $ 25 | 0.89 | 1.12 | $ 265 | $ 259 | 89% |
| 1.02C.05.08 | Applications | $ 404 | $ 348 | $ 268 | $ (57) | $ 80 | 0.86 | 1.30 | $ 632 | $ 559 | 55% |
| 1.02C.05.10 | Integration & Test | $ 107 | $ 104 | $ 86 | $ (3) | $ 18 | 0.97 | 1.21 | $ 627 | $ 627 | 17% |

- 2 persons; 0.8 FTE
- 0.5 FTE scientist, 0.3 Software Engineer
- DM-10 invoked - see DMTN-096

# Data Archive & Application Services (02C.06)

**CHARGE:** 2, 3

| WBS | DESCRIPTION | CUMULATIVE TO DATE ($k) 31 May 2019 — BUDGETED COST: WORK SCHEDULED | BUDGETED COST: WORK PERFORMED | ACTUAL: COST OF WORK PERFORMED | VARIANCE: SCHEDULE | VARIANCE: COST | PERFORMANCE: SPI | PERFORMANCE: CPI | AT COMPLETION: BAC | AT COMPLETION: EAC | AT COMPLETION: % COMP |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1.02C | Data Management Construction | $ 9,409 | $ 8,841 | $ 8,983 | $ (568) | $ (143) | 0.94 | 0.98 | $ 13,052 | $ 13,400 | 68% |
| 1.02C.06 | Science Data Archive and Application Services | $ 9,409 | $ 8,841 | $ 8,983 | $ (568) | $ (143) | 0.94 | 0.98 | $ 13,052 | $ 13,400 | 68% |
| 1.02C.06.00 | Management, Leadership & Other Costs | $ 3,377 | $ 3,377 | $ 3,493 | $ - | $ (117) | 1.00 | 0.97 | $ 4,256 | $ 4,467 | 79% |
| 1.02C.06.01 | Science Data Archive | $ 511 | $ 427 | $ 327 | $ (84) | $ 100 | 0.84 | 1.31 | $ 684 | $ 647 | 62% |
| 1.02C.06.02 | Data Access Services | $ 5,377 | $ 4,893 | $ 5,163 | $ (484) | $ (269) | 0.91 | 0.95 | $ 6,872 | $ 7,189 | 71% |
| 1.02C.06.03 | Task Framework | $ 140 | $ 140 | $ - | $ - | $ 140 | 1.00 | - | $ 1,119 | $ 980 | 12% |
| 1.02C.06.04 | Middleware Infrastructure and Toolkit | $ 4 | $ 4 | $ - | $ - | $ 4 | 1.00 | - | $ 121 | $ 117 | 3% |

- 9 persons; 6.3 FTE
- 0.9 FTE T/CAM, 8 software engineers
- New hire Christine Banek, physically located in Tucson
- Vaikunth Thukral left the project (moved to Teradata)

# LSST Data Facility (02C.07)

**CHARGE:** 2, 3

| | | CUMULATIVE TO DATE ($k) 31 May 2019 | | | | | | | | AT COMPLETION | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | BUDGETED COST | | ACTUAL | VARIANCE | | PERFORMANCE | | | | | |
| WBS | DESCRIPTION | WORK SCHEDULED | WORK PERFORMED | COST OF WORK PERFORMED | SCHEDULE | COST | SPI | CPI | BAC | EAC | % COMP |
| 1.02C | Data Management Construction | $ 18,401 | $ 17,942 | $ 17,823 | $ (460) | $ 118 | 0.98 | 1.01 | $ 42,122 | $ 42,218 | 43% |
| 1.02C.07 | Processing Control and Site Infrastructure | $ 18,401 | $ 17,942 | $ 17,823 | $ (460) | $ 118 | 0.98 | 1.01 | $ 42,122 | $ 42,218 | 43% |
| 1.02C.07.00 | Processing Control and Site Infrastructure Man | $ 3,664 | $ 3,664 | $ 3,484 | $ - | $ 181 | 1.00 | 1.05 | $ 3,664 | $ 3,484 | 100% |
| 1.02C.07.01 | Processing Control | $ 1,503 | $ 1,503 | $ 1,355 | $ - | $ 148 | 1.00 | 1.11 | $ 1,503 | $ 1,355 | 100% |
| 1.02C.07.02 | Infrastructure Services | $ 599 | $ 545 | $ 528 | $ (54) | $ 18 | 0.91 | 1.03 | $ 599 | $ 585 | 91% |
| 1.02C.07.03 | Environment and Tools | $ 179 | $ 179 | $ 192 | $ - | $ (13) | 1.00 | 0.93 | $ 179 | $ 192 | 100% |
| 1.02C.07.04 | Site Infrastructure | $ 4,216 | $ 4,216 | $ 4,386 | $ - | $ (170) | 1.00 | 0.96 | $ 4,216 | $ 4,386 | 100% |
| 1.02C.07.05 | LSST Data Facility Management, Service Archite | $ 1,561 | $ 1,542 | $ 1,484 | $ (19) | $ 58 | 0.99 | 1.04 | $ 4,048 | $ 4,007 | 38% |
| 1.02C.07.06 | LDF-offered Services | $ 604 | $ 594 | $ 512 | $ (10) | $ 83 | 0.98 | 1.16 | $ 1,137 | $ 1,059 | 52% |
| 1.02C.07.07 | Data, Compute, and IT Security Services | $ 525 | $ 464 | $ 546 | $ (61) | $ (82) | 0.88 | 0.85 | $ 1,440 | $ 1,552 | 32% |
| 1.02C.07.08 | LDF Service Software | $ 1,717 | $ 1,509 | $ 1,618 | $ (208) | $ (109) | 0.88 | 0.93 | $ 3,811 | $ 3,997 | 40% |
| 1.02C.07.09 | ITC and Facilities | $ 3,834 | $ 3,725 | $ 3,720 | $ (109) | $ 5 | 0.97 | 1.00 | $ 21,524 | $ 21,602 | 17% |

- 31 persons; 17.15 FTE
- 16 IT engineers / admins (7.85 FTE)
- 5 managers (3.25 FTE)
- 5 scientists (2.75 FTE)
- 8 software engineers (3.3 FTE)
- In addition, the Data Facility hosts 8 members of staff that are not part of the DM subsystem.

# International Comms. & Base Site (02C.08)

**CHARGE:** 2, 3

| WBS | DESCRIPTION | CUMULATIVE TO DATE ($k) 31 May 2019 | | | | | | | AT COMPLETION | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | BUDGETED COST | | ACTUAL | VARIANCE | | PERFORMANCE | | | | |
| | | WORK SCHEDULED | WORK PERFORMED | COST OF WORK PERFORMED | SCHEDULE | COST | SPI | CPI | BAC | EAC | % COMP |
| 1.02C | Data Management Construction | $ 24,400 | $ 23,826 | $ 24,059 | $ (575) | $ (233) | 0.98 | 0.99 | $ 30,487 | $ 30,686 | 78% |
| 1.02C.08 | International Communications and Base Site | $ 24,400 | $ 23,826 | $ 24,059 | $ (575) | $ (233) | 0.98 | 0.99 | $ 30,487 | $ 30,686 | 78% |
| 1.02C.08.00 | International Communications and Base Site M | $ 523 | $ 523 | $ 888 | $ - | $ (365) | 1.00 | 0.59 | $ 653 | $ 1,027 | 80% |
| 1.02C.08.01 | Base Center | $ 999 | $ 872 | $ 877 | $ (127) | $ (5) | 0.87 | 0.99 | $ 1,641 | $ 1,676 | 53% |
| 1.02C.08.02 | Chilean Data Access Center | $ - | $ - | $ 4 | $ - | $ (4) | - | - | $ - | $ 4 | 0% |
| 1.02C.08.03 | Long-Haul Networks | $ 22,878 | $ 22,430 | $ 22,290 | $ (447) | $ 140 | 0.98 | 1.01 | $ 28,194 | $ 27,979 | 80% |
| 1.04C | Telescope and Site Construction | $ 1,047 | $ 992 | $ 1,407 | $ (55) | $ (415) | 0.95 | 0.70 | $ 1,160 | $ 1,567 | 85% |
| 1.04C.12 | Utilities and Support Equipment | $ 1,047 | $ 992 | $ 1,407 | $ (55) | $ (415) | 0.95 | 0.70 | $ 1,160 | $ 1,567 | 85% |
| 1.04C.12.05 | Summit Network System | $ 1,047 | $ 992 | $ 1,407 | $ (55) | $ (415) | 0.95 | 0.70 | $ 1,160 | $ 1,567 | 85% |

- 4 persons; 1.7 FTE
- 1 T/CAM, 1 Network Architect, 1 engineers 1 technician
- Further 1.5 FTE shared with Telescope & Site subsystem
- Network Architect Ron Lambert now based in Canada
- IT System Engineer Villalobos recently departed; temporary replacement in place.

# Science Quality & Reliability Engineering (02C.10)

**CHARGE:** 2, 3

| | | CUMULATIVE TO DATE ($k) 31 May 2019 | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | BUDGETED COST | | ACTUAL | VARIANCE | | PERFORMANCE | | AT COMPLETION | | |
| WBS | DESCRIPTION | WORK SCHEDULED | WORK PERFORMED | COST OF WORK PERFORMED | SCHEDULE | COST | SPI | CPI | BAC | EAC | % COMP |
| **1.02C** | **Data Management Construction** | **$ 4,315** | **$ 4,232** | **$ 4,134** | **$ (83)** | **$ 98** | **0.98** | **1.02** | **$ 7,603** | **$ 7,512** | **56%** |
| **1.02C.01** | **System Management** | **$ 949** | **$ 949** | **$ 913** | **$ -** | **$ 36** | **1.00** | **1.04** | **$ 949** | **$ 913** | **100%** |
| 1.02C.01.02 | Science Data Quality Integration and Test | $ 949 | $ 949 | $ 913 | $ - | $ 36 | 1.00 | 1.04 | $ 949 | $ 913 | 100% |
| **1.02C.10** | **Science Quality and Reliability Engineering** | **$ 3,366** | **$ 3,282** | **$ 3,221** | **$ (83)** | **$ 61** | **0.98** | **1.02** | **$ 6,654** | **$ 6,599** | **49%** |
| 1.02C.10.01 | SQuaRE Management and Communications Lat | $ 1,613 | $ 1,613 | $ 1,623 | $ - | $ (11) | 1.00 | 0.99 | $ 2,113 | $ 2,123 | 76% |
| 1.02C.10.02 | SQuaRE Management and Communications Prc | $ 1,753 | $ 1,670 | $ 1,598 | $ (83) | $ 72 | 0.95 | 1.05 | $ 4,541 | $ 4,476 | 37% |

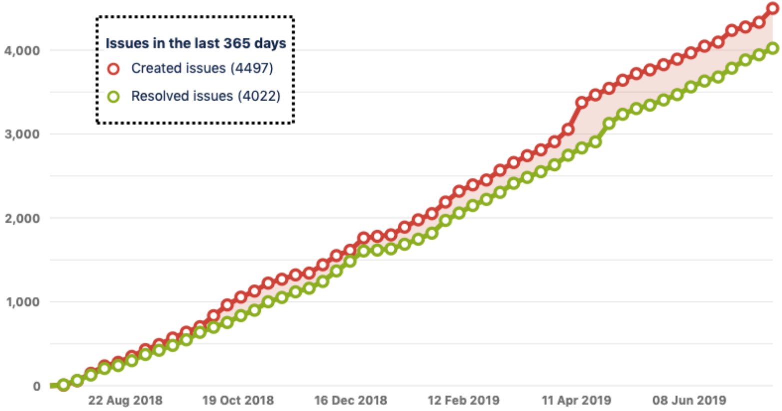
– 6 persons; 6 FTE

– 1 T/CAM, 1 Science Lead, 4 engineers

# Outline

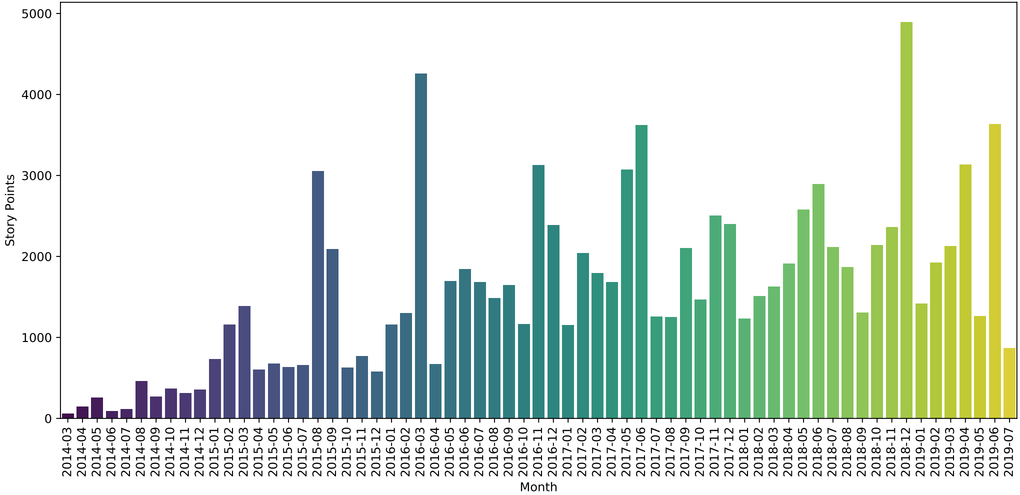
# Jira Issues: Cumulative Created vs. Resolved

CHARGE: 1, 3



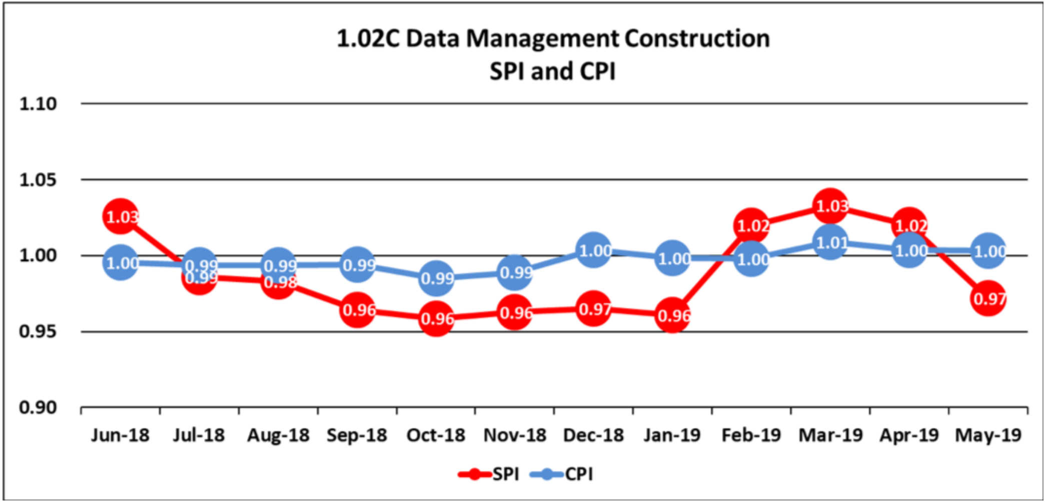
# Story Points Completed Per Month

CHARGE: 1, 3



# 1.02C (DM) Schedule & Cost Variance

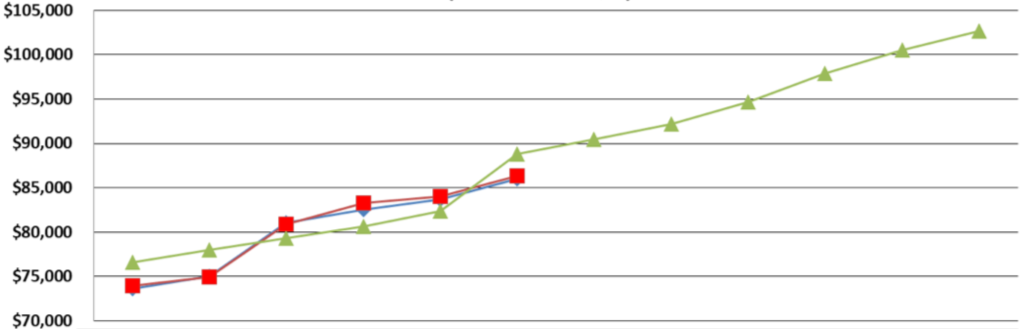
CHARGE: 2



# 1.02C (DM) Cumulative Earned Value

CHARGE: 2

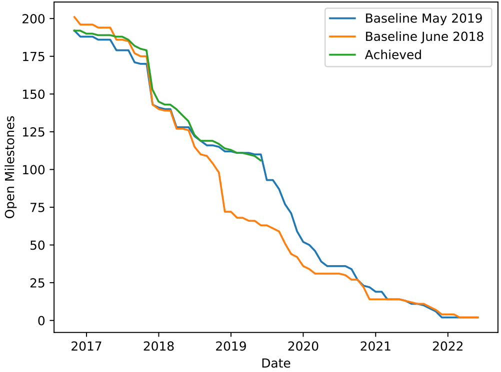
**1.02C Data Management Construction Monthly Earned Value Cumulative (all Values in $K)**




| | Dec-18 | Jan-19 | Feb-19 | Mar-19 | Apr-19 | May-19 | Jun-19 | Jul-19 | Aug-19 | Sep-19 | Oct-19 | Nov-19 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ACWP | $73,642 | $75,062 | $81,025 | $82,548 | $83,701 | $86,010 | | | | | | |
| BCWP | $73,913 | $74,940 | $80,871 | $83,290 | $84,005 | $86,292 | | | | | | |
| BCWS | $76,569 | $77,992 | $79,296 | $80,636 | $82,352 | $88,798 | $90,466 | $92,177 | $94,668 | $97,920 | $100,518 | $102,706 |

# Milestone Burndown

CHARGE: 1, 2




- All scheduled level 2 test milestones have been achieved since the last review (some a little late)
  - LDM-503-07 Camera Data Processing—DMTR-112
  - LDM-503-08 Spectrograph Data Acquisition—DMTR-121
  - LDM-503-08b Small Scale CCOB Data Access—DMTR-102
  - LDM-503-09a Science Pipelines Fall 2018 Release—DMTR-111
  - DMTC-8100-2112 Miami-Boca Raton path diverse fiber
- We are carrying only 8 missed milestones (details in breakout)

# Milestone Burndown I



- Most level 3 milestones completed on schedule.
- The following are currently late:
  - DM-DRP-8: Calibration product generation for AuxTel, due 2019-06-03: AuxTel late.
  - DM-DAX-9, DM-DAX-10, DLP-802, DLP-837, DLP-808: Middleware enhancements, due 2019-06-03; obsoleted by “generation 3” middleware.
  - DM-DAX-2: IVOA TAP async queries, due 2019-05-01: was pending staffing, now underway.
  - DM-DAX-3: Image cutout supporting IVOA SODA, due 2019-06-03: pending staffing, prototype available.
  - DM-DAX-3: Image cutout service supporting IVOA SODA protocol, due 2019-06-03: was pending staffing, now underway
  - DM-DAX-4: Metadata service supporting IVOA SIAv2 protocol, due 2019-06-03: was pending staffing, now underway

# Milestone Burndown II

**CHARGE:** 1, 2

- DM-DAX-7: Butler interface to retrieve images from data backbone, due 2019-06-03: pending backbone development
- DM-NCSA-8: Test instance of feeds to LSST mini broker in online (live stream) and offline (replaying from files) modes, due 2019-06-03: awaiting deployment of filtering service

# Outline

# Risk Management

CHARGE: 1, 3

- DM follows the procedure in LPM-20, the LSST Risk & Opportunity Management Plan, for handling risks.
- Anyone in DM can raise a risk or opportunity for assessment by the DM CCB and potential forwarding to LSST CCB.
- All risks are in the LSST Risk Register `https://jira.lsstcorp.org/projects/RM/issues`.
- There are a total of 66 (down from 84) risks currently being tracked for DM.
- Risks are assigned to WBS leads who assess and monitor them.
- See the latest project-level risk report for details.
- DM risk exposure has reduced significantly (by ∼$7M) since last year.

# Risks

CHARGE: 1

**Risk Management / Rich Filter Controller**

**Risk Management / Risks**

Exposure | Response | Dates | People

| T | LR | Key | Components | Summary | Status | RS | PWE (C |
|---|---|---|---|---|---|---|---|
| ! | < 30 DAYS | RM-773 | Data Management | Computing power required for Data Release Production exceeds estimates by large factor | ACTIVE RISK/OPPOR... | ● | 1 |
| ! | < 30 DAYS | RM-775 | Data Management | Unanticipated characteristics of real data result in poor MultiFit performance (computational) | ACTIVE RISK/OPPOR... | ● | |
| ! | < 30 DAYS | RM-723 | Data Management | Object counts exceed expectations, leading to insufficient compute | ACTIVE RISK/OPPOR... | ● | |
| ! | < 30 DAYS | RM-671 | Data Management | Computing power required for Alert Production exceeds estimates by large factor | ACTIVE RISK/OPPOR... | ● | |
| ! | < 30 DAYS | RM-752 | Data Management | Computing cost performance | ACTIVE RISK/OPPOR... | ● | |
| ! | < 30 DAYS | RM-672 | Data Management | Computing power required for Data Release Production exceeds estimates by small factor | ACTIVE RISK/OPPOR... | ● | |
| ! | < 30 DAYS | RM-673 | Data Management | Archive sizing inadequate | ACTIVE RISK/OPPOR... | ● | |
| ! | < 30 DAYS | RM-675 | Data Management | Computing power required for Alert Production exceeds estimates by small factor | ACTIVE RISK/OPPOR... | ● | |
| ! | < 30 DAYS | RM-676 | Data Management | Alert false alarm rate | ACTIVE RISK/OPPOR... | ● | |
| ! | < 30 DAYS | RM-768 | Data Management | LSST DM software architecture incompatible with de-facto community standards | ACTIVE RISK/OPPOR... | ● | |

**Risk Management / Current Filter Risk Exposure**

**66**
Issue Count

**8,379.47**
Σ Probability Weighted Exposure (Current K$)

**6,775.14**
Σ Probability Weighted Exposure (Current K$) After Mitigati...

66 filtered issues (184 hidden)

**Risk Management / Risk Review Dates**

| 0 | 1 | 67 |
|---|---|---|
| > 60 DAYS | 30-60 DAYS | < 30 DAYS |

**Issue Count** by **Last Reviewed**
68 filtered issues (859 hidden)

**Risk Management / Statistics**

# Top risks

CHARGE: 1

- RM-773: Computing power required for Data Release Production exceeds estimates; $1.3M
  - Continuing to refine our compute budget through lots of testing on precursor data.
  - Revised sizing model being prepared by the Data Facility team.
  - Amazon study, including spot pricing.
- RM-775: Unanticipated characteristics of real data result in poor MultiFit performance (computational); $0.9M
  - We expect to move away from the MultiFit approach to object characterization; a change request will be forthcoming.
  - All algorithms — not just MultiFit — are vulnerable to the "unanticpated characteristics of real data", and hence we continue to carry out extensive tests with precursor data.
- RM-723: Object counts exceed expectations, leading to insufficient compute; $0.8M
  - We are working with the Project Science Team to refine requirements around LSST processing of "crowded fields".

# Current Worries

CHARGE: 1, 3

- Impact of project security policy on DM service deployments at the Data Facility.
  - Could be avoiding by deploying to the commodity cloud instead, but there are cost implications.
- Use of U. Illinois Kerberos on Cerro Pachón.
  - One solution would be for NCOA to take on identity management and security.
- DM now is now supporting development by the Telescope & Site Software and Project IT groups.
  - It is in all our interests to see LSST succeed!
  - So far, this hasn't had a major impact on DM.
- Leaning a lot on Swinbank (DM) and Clements (T&S Software)...
  - ...but they seem to be doing fine.

# Outline

Introduction

Estimate to Complete & Current Staffing (Charge 2, 3)

Progress tracking (Charge 1, 2, 3)

Risks and Opportunities (Charge 1, 3)

Previous Recomendations (Charge 7)

Conclusion

**LIT-525 Reassess within the next year risks associated with relying on a non-path-redundant summit-to-base optical fiber network connection. . .**

. . . based on recent project experience.

- We examined all the posts in the fiber run and implemented the following improvements:
  - A single contractor was engaged for repair, improvement, and maintenance of power and fiber posts.
  - Fiber was moved to separate posts (not on same posts as power lines) in the steepest parts of the path up to Cerro Pachón.
  - Wooden posts were replaced by concrete posts in areas where both fiber and power lines are on the same post, such as in riverbeds and bends $> 20°$.

# Recs. from July 2018 DM Review II

  - Wooden posts were retained where terrain and access did not permit concrete posts.
  - Brackets connecting fibers to posts were corrected where they had been mounted improperly.

- We will continue to monitor the failure rate and maintenance status of the posts, and a diverse underground path remains a future option.

# Recs. from July 2018 DM Review III

**LIT-526 Start a process within the next 6 months to demonstrate how the Science Platform will satisfy the full suite of envisioned users and use cases. . .**

. . . by engaging the user community to evaluate the functionality of the Science Platform against those use cases and different types of LSST users.

- The Science Platform Review Final Design Review was held on 10 April 2019.
  - Reviewers included both internal and external participants.
  - See the review site for details of the committee, charge and agenda: `https://project.lsst.org/reviews/lsp-fdr/` (credentials required).
- We regard this as just the start of an ongoing process, including acting on the large number of recommendations engendered by that review and continuing to engage with stakeholders throughout the scientific community.

# Recs. from April 2019 LSP Review I

**LIT-546 Reevaluate potential user demand for the LSP (1a)**

1) While the LSP represents an extremely powerful and groundbreaking new approach to astronomical data analysis, the level of demand for its services remains unclear and there is a significant risk that demand could exceed available resources, potentially by a large margin. The committee recommends that the Project: (. . . )
a) Reevaluate potential user demand, based on recent usage of the Gaia archive and clearly document the assumptions.

– We agree with this. We will model user demand and assess how it matches to the existing DM plans for computing infrastructure resources.

**LIT-547 Reassess LSP storage and query load performance requirements (1b)**

1) While the LSP represents an extremely powerful and groundbreaking new approach to astronomical data analysis, the level of demand for its services remains unclear and there is a significant risk that demand could exceed available resources, potentially by a large margin. The committee recommends that the Project: (. . . )

b) Reassess the performance requirements for user storage and simultaneous query load in light of the new estimates of demand.

– Included in the actions taken in response to 1a (LIT-546).

# Recs. from April 2019 LSP Review III

**LIT-548 Review FTEs allocated to the long-term maintenance of the LSP (1c)**

1) While the LSP represents an extremely powerful and groundbreaking new approach to astronomical data analysis, the level of demand for its services remains unclear and there is a significant risk that demand could exceed available resources, potentially by a large margin. The committee recommends that the Project: (. . . )

c) Review the FTEs allocated to the long-term maintenance of the LSP in the light of the proposed re-evaluation of user demand.

– A focus of the LSP design and engineering is on producing a system that scales well, including in the provision of self-service facilities to cover anticipated user requests. Nevertheless, we will examine maintenance staffing in the light of any revised user demand figures.

# Recs. from April 2019 LSP Review IV

**LIT-549 Perform a cost/benefit analysis of a Cloud-based deployment of the LSP (1d)**

1) While the LSP represents an extremely powerful and groundbreaking new approach to astronomical data analysis, the level of demand for its services remains unclear and there is a significant risk that demand could exceed available resources, potentially by a large margin. The committee recommends that the Project: (. . . )

d) Perform a cost/benefit analysis of a Cloud-based deployment of the LSP (. . . )

# Recs. from April 2019 LSP Review V

– We have already performed experiments with cloud-based deployment of several of the components of the LSP, and have been in discussion with cloud providers regarding pricing. We will continue these activities.

# Recs. from April 2019 LSP Review VI

CHARGE: 7

**LIT-550 Decide on whether to adopt a Cloud-based LSP design (1d, 2a)**

1) While the LSP represents an extremely powerful and groundbreaking new approach to astronomical data analysis, the level of demand for its services remains unclear and there is a significant risk that demand could exceed available resources, potentially by a large margin. The committee recommends that the Project: (. . . )

d) (. . . ) determine whether to include a Cloud-based deployment of the LSP in the baseline design as soon as possible; and

2a) If a Cloud-based model is adopted for the LSP, the baseline plan should be revised accordingly.

# Recs. from April 2019 LSP Review VII

CHARGE: 7

– The LSP is already cloud based: we currently deploy on a private Kubernetes cloud at NCSA. We assume this recommendation means deployment on the commercial cloud, which we are working to assess.

**LIT-551 Explore ways to support the development of other DACs (1e)**

1) While the LSP represents an extremely powerful and groundbreaking new approach to astronomical data analysis, the level of demand for its services remains unclear and there is a significant risk that demand could exceed available resources, potentially by a large margin. The committee recommends that the Project: (. . . )

e) Explore ways to encourage and support the development of other DACs, ideally distributed across the world offering similar functionality, as well as DACs which may wish to serve a subset of the LSST data with an interface customized to a specific user-base.

# Recs. from April 2019 LSP Review IX

CHARGE: 7

– An international DAC policy document (LPM-251) is in preparation, and we are already in contact at a technical level with several other existing astronomical data facilities. We expect that the future evolution of international participation in LSST will include the construction of DAC-like facilities, and we already understand that our likely partners are interested in replicating the LSP environment.

**LIT-552 Plan for and provide cross-identification tables between LSST objects and external catalogs (1f)**

1) While the LSP represents an extremely powerful and groundbreaking new approach to astronomical data analysis, the level of demand for its services remains unclear and there is a significant risk that demand could exceed available resources, potentially by a large margin. The committee recommends that the Project: (. . . )

f) Provide baseline “neighbors tables” cross-identifying LSST sources/objects against other major contemporary catalogs, such as the Gaia final data release. These are likely to be required by the commissioning teams, and this will avoid the unnecessary overhead of many users attempting to perform the cross-matching independently. Engage the SCs in defining a suitable baseline.

- We will ask the DM System Science Team to consider this question over the next 9–12 months, and to sketch out a small range of alternate scenarios which the DM technical team can then evaluate for costs and benefits.
- We will be open both to the “neighbors table” approach discussed in the review, as well as to the possibility of community-contributed cross-match tables which the Project might be able to host as a service.

**LIT-553 Complete the LSP requirements and their flowdown (2)**

2) Some elements of the LSP requirements and verification/validation documents remain incomplete. The committee acknowledges legitimate reasons why the development process needed to be able to explore the best available tools, which evolve rapidly. Nevertheless, the specification of the requirements should be completed, which will enable the projects verification process to be completed and thoroughly applied. The committee feels that this should not delay LSP development.

- We address this recommendation in two parts:
  - Part 1: flowdown of, and defining V&V plans for, existing requirements. This activity is underway. We expect flowdown to be completed during summer 2019; verification plans will be completed later in the year.

# Recs. from April 2019 LSP Review XIII

CHARGE: 7

- Part 2: identification of missing requirements. This will be prioritized following the completion of part 1. We hope to have results in time to incorporate into our Spring 2020 development cycle.

**LIT-554 Document intent for finalization of LSP design (2b)**

2b) If there will be a point at which no further tools can be added (to the LSP), this should be clearly documented.

- There is a core level of LSP capabilities that the Project is funded to provide, as defined in the LSP requirement. However, the LSP design is, and is intended to be, open to future evolution.
- The use of community standard software, such as Kubernetes and Jupyter, and of community standard (i.e., IVOA) data interfaces, facilitates this sort of further evolution.
- The LSST Operations Project will be responsible for determining the course of future development of the LSP.

# Recs. from April 2019 LSP Review XV

**LIT-555 Prioritize all LSP requirements (2c)**

2c) The priority of each requirement should also be documented.

- We will do this in time for the August 2019 Joint Status Review, following the scheme used in the DMSR (LSE-61).

**LIT-556 Schedule a followup LSP design review in 2020 (2d)**

2d) The committee recommends that a program of regular reviews be adopted for the LSP throughout the lifetime of LSST, to ensure that the LSP remains responsive to users needs and software trends. At minimum, the committee recommends an additional design review in $\sim$1 year from the date of this report.

– We will plan to have a followup review in Fall 2020, which will cover:
1. completion of the requirements and V&V documents;
2. substantive fleshing out of the design for user databases and parallel/batch processing support;
3. results of the user-facing validation exercise with reprocessed HSC data which will be conducted in Spring 2020.

- Beyond that, we do not plan to have further LSP-specific reviews, but we will have an ongoing program of engagement with the user community (as discussed below in this response) and a series of substantive opportunities for user testing of the LSP in the context of access to commissioning data, notably when ComCam on-sky data are available.
- We will regularly report on the LSP status, including the results of user testing, at the project's annual status reviews.

**LIT-557 Consider up-scoping the Portal Aspect of the LSP (3a1)**

3) The Portal aspect is likely to remain an important mechanism for many users to discover, explore and exploit LSST data. Its descoping is likely to be seriously detrimental to a number of science cases (notably Solar System and time-domain science) and significantly hamper the ability of users who might prefer to use a language other than python to access and analyze LSST data. Regardless of language, some tasks benefit from GUI-like visualization tools.

a) The committee recommends that the Project explore the feasibility of up-scoping the Portal aspect, ideally including time-series plotting functions, or at a minimum, a commitment to launch and maintain it at the current level of functionality. Time-series plotting functionality should be provided within the notebook aspect, even if it has to be descoped from the Portal.

- The portal work will be reactivated closer to DR1 as stated in DMTN-096.
- One can spend infinite resources in this area... but there is little chance of an increase in funding in the current climate. The specific recommendations made by the committee will be considered as and when funding becomes available.
- We are open to in-kind contributions in this area.

# Recs. from April 2019 LSP Review XX

**LIT-558 Provide a scheme for time-series display in the Notebook Aspect (3a2)**

3a) (. . . ) Time-series plotting functionality should be provided within the notebook aspect, even if it has to be descoped from the Portal.

– We accept this recommendation and will ensure that a convenient Python API for time-series plotting is provided.

**LIT-559 Document a plan for resource allocation policy mechanisms for the LSP (4)**

The Project should clearly document their plan to ensure fair allocation of resources for a large and diverse community of scientific users, including mechanisms to enable users to temporarily expand their storage/CPU limits for clearly defined periods and a proposal mechanism for users with long-term high-usage requirements. It does not seem reasonable to ask users to submit a proposal when they need a temporary increase in resources; this should be handled via a temporary resource availability option (e.g., at least some scratch space that is cleaned up periodically) which is clearly documented. For longer term, larger or more specialized needs, this could include users purchasing additional computing resources and/or disk volumes for inclusion in the LSP resource pool, either in a local DAC or in the Cloud, and the option to apply for additional resources via competitive proposal.

– We recognize the importance of a scheme for providing temporary “scratch” resources. We will write a policy that takes this concern into account, along with defining any design details required to support it.

**LIT-560** **Consider whether the Project can accommodate computing resources donated by users to the LSP (4a)**

4a) (. . . ) For longer term, larger or more specialized needs, (a policy for LSP user resource management) could include users purchasing additional computing resources and/or disk volumes for inclusion in the LSP resource pool, either in a local DAC or in the Cloud, and the option to apply for additional resources via competitive proposal.

# Recs. from April 2019 LSP Review XXIV

CHARGE: 7

- The availability of additional (more-than-temporary) resource allocations via competitive proposal is a long-standing part of the LSST baseline. The ability for users to contribute additional resources is a feature of many large-scale HEP systems, and its potential relevance to LSST is clear. While this is beyond the current baseline, both a cloud model and an NCSA-based model can accommodate this in principle, with different degrees of complexity. The Kubernetes architecture adopted for the LSP facilitates the incorporation of user resources via the “namespace” mechanism.
- We will evaluate the technical and cost impact of this recommendation in detail and make a recommendation to LSST Project Management.

# Recs. from April 2019 LSP Review XXV

**LIT-561 Encourage each Science Collaboration to be represented in the Stack Club (5a)**

5) Science users should be introduced to the LSP at the earliest possible opportunity, to afford them time to adapt their workflow and existing analysis software to the new interfaces. (. . . )

a) This could be achieved by gradually increasing user-engagement, starting with the LSST Stack Club, in a phased program, and/or reaching out to the HSC science community as with the LSST Stack. The Science Collaborations and Project should be encouraged to ensure that each Science Collaboration is represented by at least one member in the Stack Club.

# Recs. from April 2019 LSP Review XXVI

CHARGE: 7

– We will build on the successful experience with the Stack Club and define an “LSP Club” activity. Recognizing that a primary user interest will be in working with catalog data, we will plan to ramp up this activity in early 2020 once an “LSST-like” dataset is available in the LSP from an HSC public data reprocessing.

# Recs. from April 2019 LSP Review XXVII

**LIT-562 LSP team should communicate directly with each Science Collaboration (5b)**

5) Science users should be introduced to the LSP at the earliest possible opportunity, to afford them time to adapt their workflow and existing analysis software to the new interfaces. (. . . )

b) The LSP Team should engage with each Science Collaboration directly to ensure their users are aware of the LSP capabilities and that their needs are accommodated.

– Our intent is that the primary mechanism for direct communication with Science Collaborations will be developing a relationship with a specific representative of each collaboration who will participate in the LSP Club with the larger interests of her/his whole collaboration in mind.

# Recs. from April 2019 LSP Review XXVIII

- In addition, the project will continue to make occasional presentations on the LSP's capabilities and development to representative bodies such as the PST, SAC, and SC-chairs-committee, and at the annual Project and Community Workshops.
- However we are trying to reduce cost and commissioning has priority.

**LIT-563 Construct and publicize a plan for progressive engagement of SC members with the LSP (5c)**

5) Science users should be introduced to the LSP at the earliest possible opportunity, to afford them time to adapt their workflow and existing analysis software to the new interfaces. (. . . )

c) A clear plan and timeline for user engagement should be drawn up and publicized.

– Will be answered together with 5a (LIT-561).

**LIT-564 Explore ways to involve SC (Science Collaboration) members in development of beyond-baseline LSP capabilities (5d)**

5) Science users should be introduced to the LSP at the earliest possible opportunity, to afford them time to adapt their workflow and existing analysis software to the new interfaces. (. . . )

d) If descoping is necessary, the Project should explore ways to enable a limited number of community developers to contribute to LSP functionality.

– The LSP software (and indeed the entire body of DM software) is strictly open-source and publicly hosted on Github; we are always open to pull requests from the community, assuming they meet DM engineering standards.

– As the software matures, we will provide a Contributors' Guide to aid members of the community in making valuable contributions, and will include specific examples of additional functionality that would be particularly useful.

**LIT-565 Ensure that LSP users are aware of the resources available to them (5e1)**

5) Science users should be introduced to the LSP at the earliest possible opportunity, to afford them time to adapt their workflow and existing analysis software to the new interfaces. (. . . )

e) The resources allocated to users should be clearly explained when an account is allocated to them. (. . . )

– We agree with this, and will provide documentation and an online resource query interface with this information.

**LIT-566 Devise a mechanism to identify and manage heavy users of the LSP (5e2)**

5) Science users should be introduced to the LSP at the earliest possible opportunity, to afford them time to adapt their workflow and existing analysis software to the new interfaces. (. . . )
e) (. . . ) A mechanism to identify and manage heavy users should be planned and implemented.

– We will implement such a mechanism for all resources that do not have fixed per-user limits.

# Recs. from April 2019 LSP Review XXXIV

**LIT-567 Develop a long-term data archive and software sustainability plan (6)**
6) The Project should proceed with development of an LSST data (and software) archival plan consistent with requirements in LSR and OSS to allow for long-term scientific reproducibility.

– This is not a specifically LSP issue, but deserves a project-level response.

**LIT-568 Continue with the VO-first plan for data interfaces (7)**

Clearly document, with baseline changes if necessary, that the VO-first plan is now the project baseline. . .

7) The committee supports the VO-first stances adopted by the project, since it ensures compatibility of the LSP with external services such as Vizier and with many commonly-used astronomical analysis packages. This should be continued.

- The project welcomes this advice and will continue along this path. . .
- . . . and will ensure it is properly documented!

**LIT-569 Consider parallelization technologies beyond Dask (8a)**

8a) The Dask software to enable parallelization of tasks appears to be an extremely powerful way to use the available resources with maximum efficiency and minimize computational overhead. A number of alternative parallelization packages (e.g. Parsl) are also available. These should be explored further and, if suitable, should be formally incorporated with the LSP design.

– We are not actively exploring other technologies at this time. We do plan to draw up a statement for how Dask maps onto key next-to-notebook science use cases before baselining the choice of Dask.

– We do anticipate that the Project will continue to explore new technological choices in future, but have to limit the scope we adopt for commissioning and early operations.

**LIT-570 Define a multi-image batch cutout service (8b)**

8b) The Project should formalise procedures for supporting a batch processing / bulk image cutout service, providing trivial access to imaging and relevant metadata as stored in a FITS header (e.g., gains, exposure times, WCS information).

– We are considering how how the existing SODA interface, or compatible extensions, or some alternative interface, can satisfy requests for large numbers of cutouts as a group (e.g., cutouts to accompany a full time series, or to enable a MultiFit-like analysis, or for a gallery of coadd cutouts around a large number of objects of interest).

– As our plans crystalize, we will formulate an appropriate requirement and submit it to our change control process.

**LIT-571 Ensure full-FPA visualization tools are available for commissioning (8c)**
8c) The Project should ensure a tool exists to quickly visualize the full focal plane in time for commissioning.

- This has been an ongoing issue beyond the scope of DM (and well out of scope of for LSP review).
- The Camera Team have used Firefly for this already, and have experimented with their own in-house solution.
- There is a potential in-kind offer from Brazil under discussion with the Camera and DM teams.
- A DM working group is currently charged with evaluating plans for future development within DM (LDM-702).

**LIT-572 Enable deterministic pseudo-random sampling in Qserv (9a)**

9) Although the Qserv database was outside the scope of this review, it is tightly integrated with the LSP functionality and was covered in the Team's presentations. The committee have no concerns about the impressive Qserv development effort, but note two science use cases that it would be useful for the project to track: (. . . )

a) The Project should consider including a priority-2 requirement for a mechanism to allow the random sampling of database tables in a reproducible way.

– We are considering ways in which this could be implemented as a no-cost upscope (i.e. with no additional budget.)

**LIT-573 Ensure that Qserv can support key 2- and 3-point correlation science use cases**

9) Although the Qserv database was outside the scope of this review, it is tightly integrated with the LSP functionality and was covered in the Team?s presentations. The committee have no concerns about the impressive Qserv development effort, but note two science use cases that it would be useful for the project to track: (. . . )

b) The Project should consult with stakeholders (DESC, and others) to ensure that Qserv can efficiently support standard 2- and 3-pt correlation function estimators, especially in light of the 1 arcmin margin parameter presented.

– Though an interesting idea this is a potentially a big upscope for Qserv — the Data Archive team are investigating.

# Outline

# Conclusion

- DM performance indexes have remained regular in the last year.
- We shall continue to work on delivering the software needed to make excellent LSST products.

# The END


DM and System Engineering teams at IPAC 2018

## Reference material

# Acronyms I

| Acronym | Description |
|---|---|
| ADQL | Astronomical Data Query Language |
| AIV | Assembly Integration and Verification |
| AP | Alerts Production |
| API | Application Programming Interface |
| AURA | Association of Universities for Research in Astronomy |
| AWS | Amazon Web Services |
| Alert | A packet of information for each source detected with signal-to-noise ratio > 5 in a difference image during Prompt Processing, containing measurement and characterization parameters based on the past 12 months of LSST observations plus small cutouts of the single-visit, template, and difference images, distributed via the internet |
| Alert Production | The principal component of Prompt Processing that processes and calibrates incoming images, performs Difference Image Analysis to identify DIASources and DIAObjects, packages and distributes the resulting Alerts, and runs the Moving Object Processing System |
| Archive | The repository for documents required by the NSF to be kept. These include documents related to design and development, construction, integration, test, and operations of the LSST observatory system. The archive is maintained using the enterprise content management system DocuShare, which is accessible through a link on the project website www.project.lsst.org |
| BAC | Budget At Completion |
| BDC | Base Data Center |
| BPS | Batch Production Service |
| Broker | Software which receives and redistributes Alerts, and may also perform processing such as filtering for certain characteristics, cross-matching with non-LSST catalogs, and/or light-curve classification, in order to identify and prioritize targets for follow-up and/or make scientific analyses. |
| Butler | A middleware component for persisting and retrieving image datasets (raw or processed), calibration reference data, and catalogs |
| C | Specific programming language (also called ANSI-C) |

# Acronyms II

| | |
|---|---|
| CAM | Control Account Manager |
| CCB | Change Control Board |
| CCOB | Camera Calibration Optical Bench |
| CI | Continuous Integration |
| CPI | Cost Performance Index |
| CPU | Central Processing Unit |
| Camera | The LSST subsystem responsible for the 3.2-gigapixel LSST camera, which will take more than 800 panoramic images of the sky every night. SLAC leads a consortium of Department of Energy laboratories to design and build the camera sensors, optics, electronics, cryostat, filters and filter exchange mechanism, and camera control system |
| Center | An entity managed by AURA that is responsible for execution of a federally funded project |
| Commissioning | A two-year phase at the end of the Construction project during which a technical team a) integrates the various technical components of the three subsystems; b) shows their compliance with ICDs and system-level requirements as detailed in the LSST Observatory System Specifications document (OSS, LSE-30); and c) performs science verification to show compliance with the survey performance specifications as detailed in the LSST Science Requirements Document (SRD, LPM-17) |
| Construction | The period during which LSST observatory facilities, components, hardware, and software are built, tested, integrated, and commissioned. Construction follows design and development and precedes operations. The LSST construction phase is funded through the NSF MREFC account |
| DAC | Data Access Center |
| DAQ | Data Acquisition System |
| DAX | Data Access Services |
| DB | DataBase |
| DCR | Document Change Request |
| DESC | Dark Energy Science Collaboration |
| DIA | Difference Image Analysis |

# Acronyms III

| | |
|---|---|
| DIAObject | A DIAObject is the association of DIASources, by coordinate, that have been detected with signal-to-noise ratio greater than 5 in at least one difference image. It is distinguished from a regular Object in that its brightness varies in time, and from a SSObject in that it is stationary (non-moving) |
| DIASource | A DIASource is a detection with signal-to-noise ratio greater than 5 in a difference image |
| DIMM | Differential Image Motion Monitor |
| DLP | DM Long Term Plan |
| DM | Data Management |
| DMCCB | DM Change Control Board |
| DMLT | DM Leadership Team |
| DMSR | DM System Requirements |
| DMTN | DM Technical Note |
| DMTR | Data Management Test Report |
| DOE | Department of Energy |
| DR | Data Release |
| DRP | Data Release Production |
| DWDM | Dense Wave Division Multiplex |
| Data Access Center | Part of the LSST Data Management System, the US and Chilean DACs will provide authorized access to the released LSST data products, software such as the Science Platform, and computational resources for data analysis. The US DAC also includes a service for distributing bulk data on daily and annual (Data Release) timescales to partner institutions, collaborations, and LSST Education and Public Outreach (EPO). |
| Data Backbone | The software that provides for data registration, retrieval, storage, transport, replication, and provenance capabilities that are compatible with the Data Butler. It allows data products to move between Facilities, Enclaves, and DACs by managing caches of files at each endpoint, including persistence to long-term archival storage (e.g. tape) |

# Acronyms IV

| | |
|---|---|
| Data Management | The LSST Subsystem responsible for the Data Management System (DMS), which will capture, store, catalog, and serve the LSST dataset to the scientific community and public. The DM team is responsible for the DMS architecture, applications, middleware, infrastructure, algorithms, and Observatory Network Design. DM is a distributed team working at LSST and partner institutions, with the DM Subsystem Manager located at LSST headquarters in Tucson |
| Data Management System | The computing infrastructure, middleware, and applications that process, store, and enable information extraction from the LSST dataset; the DMS will process peta-scale data volume, convert raw images into a faithful representation of the universe, and archive the results in a useful form. The infrastructure layer consists of the computing, storage, networking hardware, and system software. The middleware layer handles distributed processing, data access, user interface, and system operations services. The applications layer includes the data pipelines and the science data archives' products and services |
| Data Release | The approximately annual reprocessing of all LSST data, and the installation of the resulting data products in the LSST Data Access Centers, which marks the start of the two-year proprietary period |
| Data Release Production | An episode of (re)processing all of the accumulated LSST images, during which all output DR data products are generated. These episodes are planned to occur annually during the LSST survey, and the processing will be executed at the Archive Center. This includes Difference Imaging Analysis, generating deep Coadd Images, Source detection and association, creating Object and Solar System Object catalogs, and related metadata |
| Differential Chromatic Refraction | The refraction of incident light by Earth's atmosphere causes the apparent position of objects to be shifted, and the size of this shift depends on both the wavelength of the source and its airmass at the time of observation. DCR corrections are done as a part of DIA |
| DocuShare | The trade name for the enterprise management software used by LSST to archive and manage documents |
| Document | Any object (in any application supported by DocuShare or design archives such as PDMWorks or GIT) that supports project management or records milestones and deliverables of the LSST Project |
| EFD | Engineering Facilities Database |
| EIA | Early Integration Activity |
| ETC | Estimate To Complete |
| ETL | Extract Translate and Load |

# Acronyms V

| | |
|---|---|
| EVMS | Earned Value Management System |
| Earned Value | A measurement of how much work has been completed compared to how much was expected to have been completed at a given point in the project |
| FGCM | Forward Global Calibration Model |
| FITS | Flexible Image Transport System |
| FPA | Focal Plane Array |
| FTE | Full Time Equivalent |
| Firefly | A framework of software components written by IPAC for building web-based user interfaces to astronomical archives, through which data may be searched and retrieved, and viewed as FITS images, catalogs, and/or plots. Firefly tools will be integrated into the Science Platform |
| GUI | Graphical User Interface |
| HEP | High Energy Physics |
| HSC | Hyper Suprime-Cam |
| ID | Identifier (Identification) |
| IPAC | Infrared Processing and Analysis Center |
| ISR | Instrument Signal Removal |
| IT | Integration Test |
| IVOA | International Virtual-Observatory Alliance |
| Instrument Signature Removal | Instrument Signature Removal is a pipeline that applies calibration reference data in the course of raw data processing, to remove artifacts of the instrument or detector electronics, such as removal of overscan pixels, bias correction, and the application of a flat-field to correct for pixel-to-pixel variations in sensitivity |
| L1 | Level 1 (ambiguous could mean milestone or processing) |
| L2 | Level 2 (ambiguous could mean milestone or processing) |
| LATISS | LSST Atmospheric Transmission Imager and Slitless Spectrograph |

# Acronyms VI

| | |
|---|---|
| LCR | LSST Change Request |
| LDF | LSST Data Facility |
| LDM | LSST Data Management (handle for controlled documents) |
| LIT | LSST Issue Tracker |
| LOE | Level Of Effort |
| LPM | LSST Project Management (Document Handle) |
| LSE | LSST Systems Engineering (Document Handle) |
| LSP | LSST Science Platform |
| LSR | LSST System Requirements; LSE-29 |
| LSST | Large Synoptic Survey Telescope |
| MERRA | Modern-Era Retrospective analysis for Research and Applications |
| MOC | Multi Ordered Catalogue |
| MOPS | Moving Object Pipeline System |
| MOU | Memo Of Understanding |
| MYDB | My Database, the notion of having a local storage beside the queriable database to store either temporary tables or uploaded catalogs |
| NASA | National Aeronautics and Space Administration |
| NCOA | National Center for Optical-Infrared Astronomy |
| NCSA | National Center for Supercomputing Applications |
| NET | NETworking |
| NOAO | National Optical Astronomy Observatories (USA) |
| NSF | National Science Foundation |
| OCS | Observatory Control System |
| OODS | Observatory Operations Data Service |

# Acronyms VII

| | |
|---|---|
| OSS | Operations Support System |
| Object | In LSST nomenclature this refers to an astronomical object, such as a star, galaxy, or other physical entity. E.g., comets, asteroids are also Objects but typically called a Moving Object or a Solar System Object (SSObject). One of the DRP data products is a table of Objects detected by LSST which can be static, or change brightness or position with time |
| Offer | A response to a solicitation that, if accepted, would bind the offeror to perform the work described in resultant contract. Responses to sealed bidding are offers that are often referred to as 'bids' or 'sealed bids;' responses to a request for proposals (RFP, negotiated-type procurements) are offers often referred to as 'proposals' responses to a request for quotations (RFQ) are not offers and are generally called 'quotes' |
| Operations | The 10-year period following construction and commissioning during which the LSST Observatory conducts its survey |
| Operations Rehearsal | A data management system prototype project employing the same methods, tools, personnel, and technologies as the real system in order to introduce and validate new algorithms, functionality, and infrastructure. Previously referred to as a data challenge |
| Opportunity Management | The proactive art and science of planning, assessing, and handling future events to seek favorable impacts on project, cost, schedule, or performance to the extent possible. Opportunity management is a structured, formal, and disciplined activity focused on the necessary steps and planning actions to determine and exploit opportunities to the extent possible |
| PB | PetaByte |
| PDR | Preliminary Design Review |
| PDU | Power Distribution Unit |
| PSF | Point Spread Function |
| PST | Project Science Team |
| Parsl | Parallel Scripting Library `http://parsl-project.org/` |
| PipelineTask | A special kind of Task that can read its inputs and write its outputs using a Butler, in addition to being able to have them passed in and out directly as Python objects. PipelineTasks may be connected together dynamically and executed by a generic workflow system. PipelineTasks typically (but not always) delegate most of their work to nested regular Tasks |

# Acronyms VIII

| | |
|---|---|
| Project Manager | The person responsible for exercising leadership and oversight over the entire LSST project; he or she controls schedule, budget, and all contingency funds |
| Project Science Team | an operational unit within LSST that carries out specific scientific performance investigations as prioritized by the Director, the Project Manager, and the Project Scientist. Its membership includes key scientists on the Project who provide specific necessary expertise. The Project Science Team provides required scientific input on critical technical decisions as the project construction proceeds |
| Prompt Processing | The processing that occurs at the Archive Center on the nightly stream of raw images coming from the telescope, including Difference Imaging Analysis, Alert Production, and the Moving Object Processing System. This processing generates Prompt Data Products |
| QA | Quality Assurance |
| Qserv | Query Service, Proprietary LSST Database system |
| RM | Release Manager |
| RMS | Root-Mean-Square |
| Release | With regard to data pipelines or data products, a version that is cleared for distribution (i.e., has met QA specifications), is assigned a version identifier (e.g., 2.1), and does not evolve in the future to enable provenance |
| Review | Programmatic and/or technical audits of a given component of the project, where a preferably independent committee advises further project decisions, based on the current status and their evaluation of it. The reviews assess technical performance and maturity, as well as the compliance of the design and end product with the stated requirements and interfaces |
| Risk | The degree of exposure to an event that might happen to the detriment of a program, project, or other activity.ăIt is described by a combination of the probability that the risk event will occur and the consequence of the extent of loss from the occurrence, or impact.ăRisk is an inherent part of all activities, whether the activity is simple and small, or large and complex |
| Risk Management | The art and science of planning, assessing, and handling future events to avoid unfavorable impacts on project cost, schedule, or performance to the extent possible.ăRisk management is a structured, formal, and disciplined activity focused on the necessary steps and planning actions to determine and control risks to an acceptable level. Risk Management is an event-based management approach to managing uncertainty |
| SAC | Science Advisory Committee |

# Acronyms IX

| | |
|---|---|
| SC | Science Collaboration |
| SDSS | Sloan Digital Sky Survey |
| SLAC | No longer an acronym; formerly Stanford Linear Accelerator Center |
| SODA | Server-side Operations for Data Access |
| SPI | Schedule Performance Index |
| SPIE | the international society for optics and photonics |
| SQuaRE | Science Quality and Reliability Engineering |
| SQuaSH | Science Quality Analysis Harness |
| SST | System Science Team |
| SUIT | Science User Interface and Tools |
| Science Collaboration | An autonomous body of scientists interested in a particular area of science enabled by the LSST dataset, which through precursor studies, simulations, and algorithm development lays the groundwork for the large-scale science projects the LSST will enable. In addition to preparing their members to take full advantage of LSST early in its operations phase, the science collaborations have helped to define the system's science requirements, refine and promote the science case, and quality check design and development work |
| Science Pipelines | The library of software components and the algorithms and processing pipelines assembled from them that are being developed by DM to generate science-ready data products from LSST images. The Pipelines may be executed at scale as part of LSST Prompt or Data Release processing, or pieces of them may be used in a standalone mode or executed through the LSST Science Platform. The Science Pipelines are one component of the LSST Software Stack |
| Science Platform | A set of integrated web applications and services deployed at the LSST Data Access Centers (DACs) through which the scientific community will access, visualize, and perform next-to-the-data analysis of the LSST data products |
| Scope | The work needed to be accomplished in order to deliver the product, service, or result with the specified features and functions |
| Solar System Object | A solar system object is an astrophysical object that is identified as part of the Solar System: planets and their satellites, asteroids, comets, etc. This class of object had historically been referred to within the LSST Project as Moving Objects |

# Acronyms X

| | |
|---|---|
| Source | A single detection of an astrophysical object in an image, the characteristics for which are stored in the Source Catalog of the DRP database. The association of Sources that are non-moving lead to Objects; the association of moving Sources leads to Solar System Objects. (Note that in non-LSST usage "source" is often used for what LSST calls an Object.) |
| Source Association | The process of associating source detections on multiple images taken at different epochs, or in multiple passbands, with a single astronomical Object |
| Specification | One or more performance parameter(s) being established by a requirement that the delivered system or subsystem must meet |
| Stripe 82 | A 2.5ř wide equatorial band of sky covering roughly 300 square degrees that was observed repeatedly in 5 passbands during the course of the SDSS, In part for calibration purposes |
| Subsystem | Aăset of elements comprising a system within the larger LSST system that is responsible for a key technical deliverable of the project |
| Subsystem Scientist | The principal science advisor to a Subsystem Manager; he or she ensures that the subsystem specifications are appropriated for achieving the project's goals |
| Summit | The site on the Cerro Pachón, Chile mountaintop where the LSST observatory, support facilities, and infrastructure will be built |
| T&S | Telescope and Site |
| T/CAM | Technical/Control (or Cost) Account Manager |
| TAP | Table Access Protocol |
| TB | TeraByte |
| Task | Tasks are the basic unit of code re-use in the LSST Stack. They perform a well defined, logically contained piece of functionality. Tasks come standard with configuration, logging, processing metadata, and debugging features. For further details, see How to Write a Task in the source code documentation. Tasks can be nested, providing a natural way to structure - and configure - high level algorithms that delegate work to lower-level algorithms |
| UI | User Interface |
| US | United States |
| VCD | Verification Control Document |
| VNOC | Virtual Network Operations Center |

# Acronyms XI

| | |
|---|---|
| VO | Virtual Observatory |
| VOIP | Voice Over Internet Protocol |
| Validation | A process of confirming that the delivered system will provide its desired functionality; overall, a validation process includes the evaluation, integration, and test activities carried out at the system level to ensure that the final developed system satisfies the intent and performance of that system in operations |
| Verification | The process of evaluating the design, including hardware and software - to ensure the requirements have been met; verification (of requirements) is performed by test, analysis, inspection, and/or demonstration |
| WBS | Work Breakdown Structure |
| WCS | World Coordinate System |
| WISE | Wide-field Survey Explorer |
| algorithm | A computational implementation of a calculation or some method of processing |
| arcmin | arcminute minute of arc (unit of angle) |
| astrometry | In astronomy, the sub-discipline of astrometry concerns precision measurement of positions (at a reference epoch), and real and apparent motions of astrophysical objects. Real motion means 3-D motions of the object with respect to an inertial reference frame; apparent motions are an artifact of the motion of the Earth. Astrometry per se is sometimes confused with the act of determining a World Coordinate System (WCS), which is a functional characterization of the mapping from pixels in an image or spectrum to world coordinate such as (RA, Dec) or wavelength |
| background | In an image, the background consists of contributions from the sky (e.g., clouds or scattered moonlight), and from the telescope and camera optics, which must be distinguished from the astrophysical background. The sky and instrumental backgrounds are characterized and removed by the LSST processing software using a low-order spatial function whose coefficients are recorded in the image metadata |
| brighter-fatter effect | The common term used to refer to one of the photometric qualities of the LSST camera: sources with a higher flux have a broader PSF. This is accounted for during calibration |

| | |
|---|---|
| calibration | The process of translating signals produced by a measuring instrument such as a telescope and camera into physical units such as flux, which are used for scientific analysis. Calibration removes most of the contributions to the signal from environmental and instrumental factors, such that only the astronomical component remains |
| configuration | A task-specific set of configuration parameters, also called a 'config'. The config is read-only; once a task is constructed, the same configuration will be used to process all data. This makes the data processing more predictable: it does not depend on the order in which items of data are processed. This is distinct from arguments or options, which are allowed to vary from one task invocation to the next |
| jointcal | The jointcal package optimizes the astrometric and photometric calibrations of a set of astronomical images that cover a sky tract and were obtained as a series of visits, which may be spread out in time. The jointcal algorithms incorporates object matching both between visits and to reference star catalogs, and produces more accurate distortion and throughput models than if the astrometry and photometry were fit independently. Jointcal is a part of the Science Pipelines |
| metadata | General term for data about data, e.g., attributes of astronomical objects (e.g. images, sources, astroObjects, etc.) that are characteristics of the objects themselves, and facilitate the organization, preservation, and query of data sets. (E.g., a FITS header contains metadata) |
| metric | A measurable quantity which may be tracked. A metric has a name, description, unit, references, and tags (which are used for grouping). A metric is a scalar by definition. See also: aggregate metric, model metric, point metric |
| monitoring | In DM QA, this refers to the process of collecting, storing, aggregating and visualizing metrics |
| pipeline | A configured sequence of software tasks (Stages) to process data and generate data products. Example: Association Pipeline |
| point spread function | The point-spread function (PSF) is the distribution of intensity on a sensor (or image) originating from an unresolved point-source (i.e., a star). Often the PSF is not the same Airy shape as would be expected from a finite-aperture optical system, owing primarily to atmospheric effects and imperfections in the optical system and the detector |
| stack | a grouping, usually in layers (hence stack), of software packages and services to achieve a common goal. Often providing a higher level set of end user oriented services and tools |
| transient | A transient source is one that has been detected on a difference image, but has not been associated with either an astronomical object or a solar system body |

# References I


**[DMTR-102]**, Butler, M., 2019, *LDM-503-8b (Small Scale CCOB Data Access) Test Plan and Report*, DMTR-102, URL http://dmtr-102.lsst.io

**[DMTR-121]**, Butler, M., 2019, *LDM-503-8 (Spectrograph Data Acquisition) Test Plan and Report*, DMTR-121, URL http://dmtr-121.lsst.io

**[LSE-61]**, Dubois-Felsmann, G., Jenness, T., 2018, *LSST Data Management Subsystem Requirements*, LSE-61, URL https://ls.st/LSE-61

**[LPM-20]**, Krabbendam, V., Selvy, B., 2015, *Risk & Opportunity Management Plan*, LPM-20, URL https://ls.st/LPM-20

**[LDM-702]**, O'Mullane, W., 2019, *Image display working group charge*, LDM-702, URL https://ls.st/LDM-702

**[DMTN-096]**, O'Mullane, W., Swinbank, J., Guy, L., 2018, *Implementation and impacts of DM scope options*, DMTN-096, URL https://dmtn-096.lsst.io, LSST Data Management Technical Note

**[DMTR-111]**, Swinbank, J., 2018, *LDM-503-09a (Science pipelines fall 2018 release) Test Plan and Report*, DMTR-111, URL https://ls.st/DMTR-111

**[DMTR-112]**, Swinbank, J., 2018, *LDM-503-07 (Camera data processing) Test Plan and Report*, DMTR-112, URL https://ls.st/DMTR-112

**[LPM-251]**, William O'Mullane, M.G., Beth Willman, 2018, *Policy for Independent Data Access Centers*, LPM-251, URL https://ls.st/LPM-251